

# Additivity of nonlinear biomass equations

Bernard R. Parresol

**Abstract:** Two procedures that guarantee the property of additivity among the components of tree biomass and total tree biomass utilizing nonlinear functions are developed. Procedure 1 is a simple combination approach, and procedure 2 is based on nonlinear joint-generalized regression (nonlinear seemingly unrelated regressions) with parameter restrictions. Statistical theory is given for construction of confidence and prediction intervals for the two procedures. Specific examples using slash pine (*Pinus elliotii* Engelm. var. *elliotii*) biomass data are presented to demonstrate and clarify the methods behind nonlinear estimation, additivity, error modeling, and the formation of confidence and prediction intervals. Theoretical considerations and empirical evidence indicate procedure 2 is generally superior to procedure 1.<sup>1</sup> It is argued that modeling the error structure is preferable to using the logarithmic transformation to deal with the problem of heteroscedasticity. The techniques given are applicable to any quantity that can be disaggregated into logical components.

**Résumé :** Deux procédés qui assurent l'additivité entre la biomasse totale d'un arbre et ses composantes ont été développés en utilisant des fonctions non linéaires. Le procédé 1 est une simple approche de combinaison et le procédé 2 est basé sur la méthode non linéaire des moindres carrés unities et généraux (la méthode des régressions apparemment indépendantes et non linéaires) avec des restrictions sur les paramètres. La théorie statistique utilisée pour construire les intervalles de confiance et de prédiction est présentée pour chaque procédé. Des exemples spécifiques utilisant des données de biomasse du pin de Floride (*Pinus elliotii* Engelm. var. *elliotii*) sont présentés afin de démontrer et de clarifier les méthodes sur lesquelles reposent l'évaluation non linéaire, l'additivité, la modélisation des erreurs et la détermination des intervalles de confiance et de prédiction. Des considérations théoriques et l'évidence empirique indiquent que le procédé 2 est généralement meilleur que le procédé 1.<sup>1</sup> L'auteur soutient qu'il est préférable de modéliser la structure des erreurs plutôt que d'employer la transformation logarithmique pour solutionner le problème de l'hétéroscédacité. Les méthodes suggérées sont applicables à n'importe quelle quantité qui peut être désagrégée en composantes logiques.

[Traduit par la Rédaction]

## Introduction

A desirable feature of tree component regression equations is that the predictions for the components sum to the prediction from the total tree regression. Kozak (1970), Chiyenda and Kozak (1984), and Cunia and Briggs (1984, 1985) have discussed the problem of forcing additivity on a set of linear tree biomass functions. Parresol (1999) reviewed the three procedures one can use to force additivity of a set of linear component and total tree biomass regressions. The property of additivity assures regression functions that are consistent with each other. That is, if one tree component is part of another component, it is logical to expect the estimate of the part not to exceed the estimate of the whole. Also, if a component is defined as the sum of two sub-components, its regression estimate should equal the sum of the regression estimates of the two sub-components. Proper

inventory control depends on reliable and additive component estimates. Studies of ecosystem productivity, energy flows, and nutrient flows often break down biomass into component parts (Waring and Running 1998). In addition, economic analysis depends on products that can be aggregated and disaggregated with consistency (no discrepancy between components and total). From these arguments it is easy to recognize the importance of the additivity property.

Since publication of Parresol (1999), questions have arisen about additivity of nonlinear biomass equations. The purpose of this article is to present procedures for the additivity of nonlinear biomass equations and to demonstrate these procedures with examples. The examples help to address questions relating to the choice of proper model and to choosing estimates that are statistically reasonable (i.e., estimates that have statistical properties such as unbiasedness, consistency, and minimum variance). In this article I formalize conditions that ensure that predicted values calculated from nonlinear component biomass regression equations add up to those obtained from a corresponding nonlinear total biomass regression equation. I present theory on construction of confidence intervals and prediction intervals because current software packages will not calculate such quantities for the procedures presented here. It should be emphasized that the procedures given in this article are applicable to biomass studies in general, be they of a plant or animal nature. In fact, any dependent regression variable (such as time, energy, volume, and distance) that can be disaggregated into

Received March 28, 2000. Accepted December 13, 2000.  
Published on the NRC Research Press Web site on  
May 17, 2001.

**B.R. Parresol.** USDA Forest Service, Southern Research  
Station, P.O. Box 2680, Asheville, NC 28802, U.S.A.  
(e-mail: bparresol@fs.fed.us).

<sup>1</sup>A SAS program that implements procedure 2 is available  
from the author or may be purchased from the Depository of  
Unpublished Data, CISTI, National Research Council of  
Canada, Ottawa, ON K1A 0S2. No. 3274.

meaningful components can utilize these procedures to guarantee the property of additivity and to estimate the reliability of the aggregated total.

### Biomass modeling preliminaries

Tree biomass is normally estimated through the use of regression. Trees are chosen through an appropriate selection procedure for destructive sampling, and the weights or mass of the components of each tree are determined and related by regression to one or more dimensions of the standing tree. The tree is normally separated into three aboveground components: (i) bole or main stem, (ii) bole bark, and (iii) crown (branches and foliage). Occasionally a fourth component, belowground biomass, which is the stump and major roots within a fixed distance, is considered. Other tree component schemes are possible.

The process of physically collecting biomass data can be very labor intensive. The various tree components, as determined by the scheme used, can be measured directly as soon as they are separated from the tree. The only possible error may be due to faulty measurement instruments or methods. However, direct measurement may be too expensive and time consuming, so components are often subsampled. Small samples are selected from the tree component by some usually random procedure. Green and oven-dry masses of these samples are determined in the laboratory, and the results are used to estimate the entire tree component. Briggs et al. (1987) and Kleinn and Pelz (1987) describe the use of stratified subsampling and subsequent application of ratio-type estimators. Parresol (1999) gives sample calculations for the stratified ratio estimator using data from four sweetgum (*Liquidambar styraciflua* L.) trees. Valentine et al. (1984) and Gregoire et al. (1995) describe two procedures, randomized branch sampling and importance sampling, for selecting sample paths to obtain unbiased estimates of the biomass content of a tree.

Subsampling to estimate the component biomass produces an error of measurement of the component. What effect does this have on our ability to construct and estimate biomass models? In a linear context, the data-generating process is presumed to be

$$y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$$

$$y = y^* + v$$

$$v_i \sim \text{iid}(0, \sigma_v^2)$$

where  $y^*$  is the unobservable true dependent variable,  $y$  is what is observed (i.e., the sampling estimate of the component biomass), and the errors  $v_i$  are independent and identically distributed (iid) random variables with mean 0 and variance  $\sigma_v^2$ . We assume that  $\varepsilon$  and  $v$  are independent and that  $y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$  satisfies the classical assumptions (e.g., uncorrelated errors, nonstochastic regressor vector  $\mathbf{x}$ ) (see Greene 1999, Section 6.3). Given this, we have

$$y + v = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$$

$$y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon - v$$

$$= \mathbf{x}'\boldsymbol{\beta} + \omega$$

$$\omega_i \sim \text{iid}(0, \sigma_\varepsilon^2 + \sigma_v^2)$$

As long as  $v$  is uncorrelated with  $\mathbf{x}$ , this model satisfies the classical assumptions and can be estimated by least squares. The same principles apply to nonlinear models. Hence, this type of measurement error is not a problem.

Researchers have used a variety of regression models for estimating total-tree and tree-component biomass. Reviews of biomass studies (e.g., Pardé 1980; Baldwin 1987; Clark 1987; Pelz 1987; Parresol 1999) indicate that prediction equations generally have been developed utilizing one of the following three forms: (i) linear additive error, (ii) nonlinear additive error, and (iii) nonlinear multiplicative error. The linear additive error model and the nonlinear multiplicative error model (log transformed) can be fitted by standard linear least squares estimation procedures. The nonlinear additive error model results in nonlinear regression equations that require use of iterative procedures for parameter estimation. This model can be written as

$$[1] \quad y_t = f(\mathbf{x}_t, \boldsymbol{\beta}) + \varepsilon_t$$

where  $y_t$  is (total or component) biomass,  $\mathbf{x}_t$  is an ( $N \times 1$ ) nonstochastic vector of tree dimension variables,  $\boldsymbol{\beta}$  is a ( $K \times 1$ ) parameter vector,  $\varepsilon_t$  is a random error, and  $t$  represents the  $t$ th observation ( $t = 1, 2, \dots, T$ ). Note that, unlike linear specifications, the number of parameters,  $K$ , and the number of independent variables,  $N$ , do not necessarily coincide in nonlinear models. Commonly used tree dimension variables are diameter at breast height ( $D$ ),  $D^2$ , total height ( $H$ ),  $D^2H$ , age, live crown length (LCL), diameter at the base of the live crown, and sapwood area (active conducting tissue) measured at various heights in the stem.

Normally, biomass data exhibit heteroscedasticity; that is, the error variance is not constant over all observations. If eq. 1 is fitted to such data then weighted analysis, typically involving additional parameters, is necessary to achieve minimum variance parameter estimates (assuming all other regression assumptions are met; e.g., uncorrelated errors). A nonlinear statistical model consists jointly of a part that specifies the mean ( $f(\mathbf{x}_t, \boldsymbol{\beta})$ ) and a part describing variation around the mean, and the latter may well need more than one parameter  $\sigma^2$  to be adequate.

### Estimated generalized nonlinear estimation

The estimated generalized linear least squares estimator (also known as the weighted linear least squares estimator) is  $\mathbf{b} = (\mathbf{X}'\boldsymbol{\Psi}(\boldsymbol{\theta})^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Psi}(\boldsymbol{\theta})^{-1}\mathbf{y}$ , where  $\mathbf{W}(0)$  is a diagonal matrix of weights dependent on a fixed number (say  $P$ ) of parameters denoted by the ( $P \times 1$ ) vector  $\boldsymbol{\theta}$ . The dimension of  $\boldsymbol{\theta}$ , and the precise way in which  $\boldsymbol{\Psi}$  depends on  $\boldsymbol{\theta}$ , relies on what assumptions are made about the error process. Parresol (1993) shows a structure for  $\boldsymbol{\Psi}$  when  $\varepsilon$  is heteroscedastic and he provides details on the estimation of  $\boldsymbol{\theta}$  from least squares residuals ( $e_i$ ) for regression models with multiplicative heteroscedasticity.

### Deriving weights

As Parresol (1993) discusses, it often occurs that the error variance is functionally related to one or more of the model explanatory variables. If we assume each  $\sigma_i^2$  is an exponential function of  $P$  explanatory variables multiplied by the scale factor  $\sigma^2$ , then

$$[2] \quad E[\epsilon_t^2] = \sigma_t^2 = \sigma^2 \exp(\mathbf{g}_t' \boldsymbol{\theta}), \quad t = 1, 2, \dots, T$$

where  $\mathbf{g}_t' = [g_{t1}, g_{t2}, \dots, g_{tP}]$  is a  $(1 \times P)$  vector containing the  $t$ th observation on  $P$  nonstochastic variables and  $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_P]'$  is a  $(P \times 1)$  vector of unknown parameters. The  $g$ s could be identical to, or functions of, the  $x$ s (the independent variables in eq. 1). The relevant  $g$ s may be obvious, and experience or past work may suggest the proper variables. Often graphical analyses of the data and (or) residuals will reveal the appropriate  $g$ s. Based on eq. 2, the error structure (variance-covariance matrix) can be written as follows:

$$[3] \quad \sigma^2 \Psi(\boldsymbol{\theta}) = \sigma^2 \begin{bmatrix} \exp(\mathbf{g}_1' \boldsymbol{\theta}) & & & \\ & \exp(\mathbf{g}_2' \boldsymbol{\theta}) & & \\ & & \ddots & \\ & & & \exp(\mathbf{g}_T' \boldsymbol{\theta}) \end{bmatrix}$$

where all of the off-diagonal elements are zero. If we use the squares of the least squares residuals as representative of  $\sigma_t^2$  in eq. 2 and take the natural logarithm of both sides of the equation, we obtain the linear model

$$[4] \quad \ln \epsilon_t^2 = \ln \sigma^2 + \mathbf{g}_t' \boldsymbol{\theta} + v_t$$

where  $v_t$  is model error and  $\ln \sigma^2$  is the constant or intercept term. Harvey (1976) showed that the  $v_t$ 's satisfy the ordinary least squares (OLS) assumptions of homoscedasticity and no autocorrelation, that is, OLS regression will provide consistent estimates of the  $\theta$ s in eq. 4. Replacing the unknown parameter vector  $\boldsymbol{\theta}$  with the consistently estimated vector,  $\hat{\boldsymbol{\theta}}$ , the following weight matrix (ensuing from eq. 3) results:

$$[5] \quad \Psi(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \exp(\mathbf{g}_1' \hat{\boldsymbol{\theta}}) & & & \\ & \exp(\mathbf{g}_2' \hat{\boldsymbol{\theta}}) & & \\ & & \ddots & \\ & & & \exp(\mathbf{g}_T' \hat{\boldsymbol{\theta}}) \end{bmatrix}$$

where  $\exp(\mathbf{g}_t' \hat{\boldsymbol{\theta}})$  is the weight factor for the  $t$ th observation.

### Computational methods

In nonlinear estimation we replace the design matrix  $\mathbf{X}$  with the partial derivatives matrix  $\mathbf{Z}(\boldsymbol{\beta})$  defined as

$$[6] \quad \mathbf{Z}(\boldsymbol{\beta}) = \frac{\partial \mathbf{f}(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \begin{bmatrix} \frac{\partial f(\mathbf{x}_1, \boldsymbol{\beta})}{\partial \beta_1} & \dots & \frac{\partial f(\mathbf{x}_1, \boldsymbol{\beta})}{\partial \beta_K} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{x}_T, \boldsymbol{\beta})}{\partial \beta_1} & \dots & \frac{\partial f(\mathbf{x}_T, \boldsymbol{\beta})}{\partial \beta_K} \end{bmatrix}$$

When this matrix of derivatives is evaluated at a particular value for  $\boldsymbol{\beta}$ , say  $\boldsymbol{\beta}_1$ , it will be written as  $\mathbf{Z}(\boldsymbol{\beta}_1)$ . The generalized nonlinear least squares estimate of the vector  $\boldsymbol{\beta}$  is that value of  $\boldsymbol{\beta}$  that minimizes the residual sum of squares:

$$171 \quad S(\boldsymbol{\beta}) = \boldsymbol{\epsilon}' \Psi^{-1} \boldsymbol{\epsilon} = [\mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\beta})]' \Psi^{-1} [\mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\beta})]$$

Many iterative algorithms are possible for minimizing the expression in eq. 7. Most have the general form

$$[8] \quad \boldsymbol{\beta}_{n+1} = \boldsymbol{\beta}_n - l_n \mathbf{P}_n \boldsymbol{\gamma}_n$$

where  $\boldsymbol{\gamma}_n = \partial S / \partial \boldsymbol{\beta}|_{\boldsymbol{\beta}}$  is the gradient vector,  $\mathbf{P}_n$  is a positive definite matrix known as the direction matrix, and  $l_n$  is a positive number known as the step length. The iterations start with an initial guess  $\boldsymbol{\beta}_0$  and proceed ( $n = 0, 1, 2, \dots$ ) until some predefined convergence criteria are met (e.g.,  $\boldsymbol{\gamma}_n \approx 0$ ) or the predefined maximum number of iterations are exhausted (failure to converge). What distinguishes alternative algorithms is the definition of  $\mathbf{P}_n$ . The Gauss-Newton algorithm is defined by  $\mathbf{P}_n = [\mathbf{Z}(\boldsymbol{\beta}_n)' \Psi(\boldsymbol{\theta})^{-1} \mathbf{Z}(\boldsymbol{\beta}_n)]^{-1}$ . If the starting vector is reasonable and the direction matrix is not ill-conditioned,<sup>3</sup> this algorithm converges quickly. The steepest descent algorithm replaces  $\mathbf{P}_n$  with the scalar value 1. The steepest descent method may converge very slowly and, therefore, is not generally recommended. It is sometimes useful when the initial values are poor. Another algorithm highly favored is that of Marquardt defined by  $\mathbf{P}_n = [\mathbf{Z}(\boldsymbol{\beta}_n)' \Psi(\boldsymbol{\theta})^{-1} \mathbf{Z}(\boldsymbol{\beta}_n) + \lambda_n \text{diag}(\mathbf{Z}(\boldsymbol{\beta}_n)' \Psi(\boldsymbol{\theta})^{-1} \mathbf{Z}(\boldsymbol{\beta}_n))]^{-1}$ . The Marquardt algorithm is a compromise between Gauss-Newton and steepest descent. As  $\lambda_n \rightarrow 0$ , the direction approaches Gauss-Newton; as  $\lambda_n \rightarrow \infty$ , the direction approaches steepest descent. This algorithm is useful when the parameter estimates are highly correlated. The Marquardt algorithm is recommended by many authors (e.g., Ratkowsky 1983; Draper and Smith 1998). A description of these and other algorithms and their direction matrices can be found in Judge et al. (1985, Appendix B) and Seber and Wild (1989). See Gallant (1987) or Seber and Wild (1989) for details and considerations surrounding  $l_n$ ,  $\lambda_n$ , and the convergence criteria.

When the process has converged we obtain the estimated generalized nonlinear least squares (EGNLS) estimate  $\mathbf{b}$ . It should be mentioned that it is possible for eq. 7 to have several local extrema and that convergence may not be to a global minimum. Thus, it is a good idea to try several starting vectors in the iteration function (eq. 8) to ensure that the minimum obtained is the global one. Occasionally one may be faced with a nearly flat gradient resulting in slow convergence and premature termination of the iterations (loss of accuracy). Trying different algorithms will often highlight this condition so that one can fine tune the convergence criterion to obtain the desired degree of accuracy. Finally, it should be noted that in a nonlinear setting the iterative algorithm can break down. In a linear setting, a low  $R^2$  or other diagnostic may suggest that the model and data are mismatched, but as long as the full rank condition is met by the regressor matrix, a linear regression can always be computed. This is not the case with nonlinear regression. A convergence failure may indicate that the model is not appropriate for the body of data. For a discussion of these and other issues surrounding optimization, see Greene (1999, section 5.5).

### Hypothesis testing and interval estimation

Under appropriate conditions the EGNLS estimate  $\mathbf{b}$  will

<sup>3</sup>Let  $\hat{\theta}_T$  be an estimator of  $\theta$  based on a sample of size  $T$ . Then  $\hat{\theta}_T$  is a consistent estimator of  $\theta$  if  $\lim_{T \rightarrow \infty} P(|\hat{\theta}_T - \theta| < \delta) = 1$ , where  $\delta$  is an arbitrarily small positive number.  $\theta_T$  is said to converge in probability to 0.

<sup>4</sup>A matrix is said to be ill conditioned when its determinant, relative to its elements, is quite small.

be approximately normally distributed with mean  $\beta$  and a variance-covariance matrix that is consistently estimated by

$$[9a] \quad s^2(\mathbf{b}) = \hat{\sigma}^2 [\mathbf{Z}(\mathbf{b})' \Psi(\hat{\theta})^{-1} \mathbf{Z}(\mathbf{b})]^{-1}$$

where the scalar  $\hat{\sigma}^2$  is the regression mean squared error (eq. 7 divided by degrees-of-freedom):

$$[9b] \quad \hat{\sigma}^2 = \frac{S(\mathbf{b})}{T - K} = \frac{[\mathbf{y} - \mathbf{f}(\mathbf{X}, \mathbf{b})]' \Psi(\hat{\theta})^{-1} [\mathbf{y} - \mathbf{f}(\mathbf{X}, \mathbf{b})]}{T - K}$$

This information can be used to form hypothesis tests and interval estimates on  $\mathbf{b}$  in an analogous manner to linear least squares. For a confidence interval on the predicted mean value  $\hat{y}_t$ , the appropriate variance estimate is

$$[10] \quad \widehat{\text{var}}(\hat{y}_t) = \mathbf{z}(\mathbf{b})_t' s^2(\mathbf{b}) \mathbf{z}(\mathbf{b})_t$$

where  $\mathbf{z}(\mathbf{b})_t$  is the  $t$ th row of  $\mathbf{Z}(\mathbf{b})$  (see eq. 6). For a prediction interval on an individual (new) outcome drawn from the distribution of  $y_t$ , the variance is

$$[11] \quad \widehat{\text{var}}(\hat{y}_{t(\text{new})}) = \hat{\sigma}^2 \psi_t(\hat{\theta}) + \mathbf{z}(\mathbf{b})_t' s^2(\mathbf{b}) \mathbf{z}(\mathbf{b})_t$$

where  $\psi_t(\hat{\theta})$  is the  $t$ th diagonal element of the weight matrix  $\Psi(\hat{\theta})$ , or put another way, it is the value of the weight function ( $\exp(\mathbf{g}'_t \hat{\theta})$ ) at observation  $t$ . For further details on nonlinear regression see Gallant (1987), Bates and Watts (1988), Seber and Wild (1989), and Greene (1999, chapter 10).

## Procedures for additivity of nonlinear biomass equations

There are three procedures for forcing additivity of a set of linear tree biomass functions (see the review by Parresol 1999) but only two for nonlinear models, depending on how the separate components are aggregated. In this section, subscripts refer to tree biomass components (crown, bole, etc.). In procedure 1, the total biomass regression function is defined as the sum of the separately calculated best regression functions of the biomass of its  $c$  components:

$$[12] \quad \begin{aligned} \hat{y}_1 &= f_1(\mathbf{x}_1, \mathbf{b}_1) \\ \hat{y}_2 &= f_2(\mathbf{x}_2, \mathbf{b}_2) \\ &\vdots \\ \hat{y}_c &= f_c(\mathbf{x}_c, \mathbf{b}_c) \\ \hat{y}_{\text{total}} &= \hat{y}_1 + \hat{y}_2 + \dots + \hat{y}_c \end{aligned}$$

Reliability (i.e., confidence intervals) of the total biomass prediction can be determined from variance properties of linear combinations:

$$[13a] \quad \text{var}(\hat{y}_{\text{total}}) = \sum_{i=1}^c \text{var}(\hat{y}_i) + 2 \sum_{i < j} \text{cov}(\hat{y}_i, \hat{y}_j)$$

where

$$[13b] \quad \text{cov}(\hat{y}_i, \hat{y}_j) = \hat{\rho}_{y_i, y_j} \sqrt{\text{var}(\hat{y}_i) \text{var}(\hat{y}_j)}$$

$\hat{\rho}_{y_i, y_j}$  = estimated correlation between  $y_i$  and  $y_j$

Procedure 2 is more general and flexible than procedure 1 and more difficult to employ. Statistical dependencies (i.e., contemporaneous correlations) among sample data are accounted for using nonlinear joint-generalized least squares regression, also known as nonlinear seemingly unrelated regressions (NSUR). A set of nonlinear regression functions are specified such that (i) each component regression contains its own independent variables, and the total-tree regression is a function of all independent variables used; (ii) each regression can use its own weight function; and (iii) the additivity is insured by setting constraints (i.e., restrictions) on the regression coefficients. The structural equations for the system of nonlinear models of biomass additivity can be specified as

$$[14] \quad \begin{aligned} y_1 &= \mathbf{f}_1(\mathbf{X}_1, \beta_1) + \epsilon_1 \\ y_2 &= \mathbf{f}_2(\mathbf{X}_2, \beta_2) + \epsilon_2 \\ &\vdots \\ y_c &= \mathbf{f}_c(\mathbf{X}_c, \beta_c) + \epsilon_c \\ y_{\text{total}} &= \mathbf{f}_{\text{total}}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_c, \beta_1, \beta_2, \dots, \beta_c) + \epsilon_{\text{total}} \end{aligned}$$

When the stochastic properties of the error vectors are specified along with the coefficient restrictions, the structural equations become a statistical model for efficient parameter estimation and reliable prediction intervals. The procedure 2 or NSUR method is usually preferable to procedure 1. If disturbances or errors in the different equations are correlated (the normal situation for biomass models), then procedure 2 (formulation in eq. 14) is superior to procedure 1 (formulation in eq. 12), because NSUR takes into account the contemporaneous correlations and results in lower variance. Two comprehensive references covering NSUR are Gallant (1987) and Srivastava and Giles (1987).

## Example using procedure 1

Consider the sample of 40 slash pine (*Pinus elliotii* Engelm. var. *elliotii*) trees in Table 1. Trees for destructive sampling were selected from unthinned plantation growth and yield plots scattered throughout the state of Louisiana, U.S.A. Trees were felled at a 0.15-m stump height, separated into components of bolewood, bole bark, and crown (branches and foliage) and weighed in the field. The 40 trees given in Table 1 are a subset of a larger data set from a biomass study by Lohrey (1984), which are used here for illustrative purposes. Tree trunks display shapes that generally fall between the frustum of a cone and the frustum of a paraboloid (Husch et al. 1982). From a conceptual standpoint, volume or mass of the main stem can be obtained as a solid-of-revolution by rotating a curve of the general form,  $Y = C(X)^{0.5}$ , around the  $X$  axis. As the exponent  $r$  changes, different solids are produced. This leads to expressions having the functional form  $Y = f(D, H)$ . From this starting point, I tried functional forms including the variables LCL and age. I selected the best biomass component equations based on scatterplots of the data and the minimum of the Akaike information criterion (AIC)" (Borowiak 1989):

<sup>4</sup>The AIC is widely used for model discrimination and is computed as  $\text{AIC}_i = \ln(\mathbf{e}'_i \mathbf{e}_i / T) + 2K_i / T$ , where  $\mathbf{e}_i$  is the residual vector of the  $i$ th alternative equation,  $K_i$  is the number of coefficients, and  $T$  is the number of observations.

**Table 1.** Green mass data for slash pine trees from the state of Louisiana, U.S.A.

Tree	$D^a$ (cm)	$H^b$ (m)	LCL <sup>c</sup> (m)	Age (years)	Green mass (kg)			
					Wood	Bark	Crown	Tree
1	5.6	7.9	2.1	21	6.5	2.3	1.0	9.8
2	6.4	8.5	1.2	21	1.4	2.6	2.1	12.1
3	8.1	10.7	2.7	20	17.6	4.5	2.3	24.4
4	8.4	11.3	3.4	21	18.5	4.3	4.2	27.0
5	9.1	11.0	4.3	21	22.6	5.4	5.6	33.6
6	9.9	13.1	3.4	21	30.6	7.4	5.5	43.5
7	10.4	14.3	5.5	32	32.9	6.7	6.4	46.0
8	11.2	14.6	4.0	19	40.6	9.3	6.2	56.1
9	11.7	14.3	4.0	21	46.0	10.7	1.7	64.4
10	12.2	14.9	6.1	21	51.6	13.1	6.1	70.8
11	11.9	16.8	4.6	32	60.4	10.1	5.4	75.9
12	13.2	13.7	4.6	20	62.8	15.2	10.7	88.7
13	12.2	15.8	5.5	19	67.5	12.9	15.3	95.7
14	13.7	18.0	2.4	32	81.2	12.5	8.7	102.4
15	14.2	16.5	6.4	19	94.3	18.2	11.2	123.7
16	15.0	20.1	3.0	32	123.4	16.5	1.7	147.6
17	15.7	16.8	4.9	21	107.3	21.5	19.7	148.5
18	16.5	17.1	4.9	20	123.8	22.1	28.9	174.8
19	16.5	17.1	4.0	21	151.6	24.6	16.8	193.0
20	19.6	13.7	6.7	16	140.4	25.1	46.2	211.7
21	17.5	19.2	7.6	19	170.4	27.4	16.8	214.6
22	17.8	18.3	6.4	21	169.6	31.7	24.0	225.3
23	18.5	17.7	6.4	21	160.3	36.9	47.5	244.7
24	19.6	19.8	7.6	19	199.8	38.7	19.7	258.2
25	18.5	22.9	4.0	32	231.6	29.6	24.6	285.8
26	19.8	18.6	6.7	21	217.9	33.9	45.8	297.6
27	20.6	17.4	5.8	21	216.0	32.6	61.2	309.8
28	21.6	17.7	8.2	20	200.6	40.2	75.4	316.2
29	19.8	18.9	7.3	19	217.5	38.5	62.0	318.0
30	22.9	19.8	8.5	19	314.8	43.1	43.2	401.1
31	23.6	18.3	7.9	20	287.1	63.4	51.7	402.2
32	23.1	18.9	7.9	21	290.9	44.3	76.7	411.9
33	24.1	21.3	9.4	21	320.1	50.6	15.6	446.3
34	26.4	19.2	7.3	19	308.6	65.7	116.0	490.3
35	24.6	25.0	5.8	32	403.0	49.8	69.8	522.6
36	25.1	19.8	8.5	21	390.4	48.8	83.5	522.7
37	29.0	20.4	8.5	20	445.2	60.4	88.0	593.6
38	28.4	26.8	1.3	32	736.4	84.0	79.9	900.3
39	31.8	27.4	8.2	32	110.9	93.8	170.2	1034.9
40	33.0	27.7	10.4	45	921.3	108.0	169.2	1198.5

<sup>a</sup>Diameter at breast height<sup>b</sup>Tree height.<sup>c</sup>Live crown length.

$$\begin{aligned}
 \hat{y}_{\text{wood}} &= b_1(D^2H)^{b_2} \\
 [15] \quad \hat{y}_{\text{bark}} &= b_1D^{b_2} \\
 \hat{y}_{\text{crown}} &= b_1D^{b_2}H^{b_3}
 \end{aligned}$$

where  $D$  is diameter at breast height and  $H$  is total tree height. Starting with a weight of unity for each observation, nonlinear least squares estimates of the coefficients in eq. 15 were obtained using the iteration function in eq. 8 (Gauss–Newton algorithm in PROC NLIN; SAS Institute Inc. 1989). Initial values ( $\beta_0$ ) to begin the iterations were obtained by log transforming the equations and applying OLS. Scatterplots of

the residuals (Fig. 1) revealed significant heteroscedasticity (as expected). An error model of the form shown in eq. 4 was fitted to the residuals to obtain  $W(0)$ . This was done for each equation in eq. 15. The coefficients for each equation were then re-estimated using the iteration function in eq. 8 with its estimated weight matrix  $W(0)$ . Table 2 lists the coefficients along with their standard errors, asymptotic confidence intervals, weight functions, and fit statistics. From examining Table 2 it can be seen that weighting reduced standard errors on five of the seven coefficients. More importantly, the confidence intervals on the coefficients for the crown biomass model indicate nonsignificance (at the  $\alpha =$

0.05 level) for two of the parameters if the model is not weighted; however, under weighted least squares the parameters  $\beta_1$  and  $\beta_3$  are significant and their importance in the model is established.

Under procedure 1 for additivity, total tree biomass is simply the sum of the components. For example, using the weighted-fit coefficients in Table 2 and the set of equations in eq. 15, a tree with  $D = 20$  cm and  $H = 17$  m gives the following:  $\hat{y}_{\text{wood}} = 186.4$  kg,  $\hat{y}_{\text{bark}} = 34.7$  kg, and  $\hat{y}_{\text{crown}} = 47.0$  kg; therefore,  $\hat{y}_{\text{total}} = 186.4 + 34.7 + 47.0 = 268.1$  kg. The error for each mean component prediction is computed using eq. 10. For  $\hat{y}_{\text{wood}}$  the partial derivatives vector  $\mathbf{z}(\mathbf{b})'$  is characterized by (recall that  $\mathbf{z}(\mathbf{b})' = \partial f(\mathbf{x}, \mathbf{b}) / \partial \mathbf{b}'$ )

$$[(D^2 H)^{b_2} \quad b_1 (D^2 H)^{b_2} \quad \ln(D^2 H)]$$

Thus, for  $D = 20$  cm and  $H = 17$  m and using the weighted-fit coefficients for wood biomass in Table 2, we obtain

$$\mathbf{z}(\mathbf{b})' = [(6800)^{1.0585} \quad 0.016 \quad 363 \quad (6800)^{1.0585} \ln(6800)] = [1 \quad 1 \quad 394.365 \quad 1645.3577]$$

The variance-covariance matrix of  $\mathbf{b}$  is computed from eq. 9a and for the EGNLS wood biomass regression is

$$\mathbf{s}^2(\mathbf{b}) = \begin{bmatrix} & & & \mathbf{x} \\ & & & \\ & & & \\ & & & \end{bmatrix} = \begin{bmatrix} & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$

Therefore, the variance of  $\hat{y}_{\text{wood}}$  from eq. 10 is

$$\widehat{\text{var}}(\hat{y}_{\text{wood}}) = [1 \quad 1394.365 \quad 1645.3577] \begin{bmatrix} 3.3004 \times 10^{-6} & -2.3577 \times 10^{-5} \\ -2.3577 \times 10^{-5} & 0.0001714 \end{bmatrix} \begin{bmatrix} 11394.365 \\ 1645.3577 \end{bmatrix} = 8.46 \text{ kg}^2$$

For the other two components we obtain  $\widehat{\text{var}}(\hat{y}_{\text{bark}}) = 0.53 \text{ kg}^2$  and  $\widehat{\text{var}}(\hat{y}_{\text{crown}}) = 10.34 \text{ kg}^2$ . The correlations (computed using PROC CORR: SAS Institute Inc. 1990) between the weighted biomass components are

$$\hat{\rho}_{\hat{y}_{\text{wood}} \hat{y}_{\text{bark}}} = 0.19, \hat{\rho}_{\hat{y}_{\text{wood}} \hat{y}_{\text{crown}}} = 0.01, \text{ and } \hat{\rho}_{\hat{y}_{\text{bark}} \hat{y}_{\text{crown}}} = 0.29$$

therefore, using eq. 13 we obtain

$$\begin{aligned} \widehat{\text{var}}(\hat{y}_{\text{total}}) &= 8.46 + 0.53 + 10.34 + 2 \times 0.19 \times \sqrt{8.46 \times 0.53} + 2 \times 0.01 \times \sqrt{8.46 \times 10.34} + 2 \times 0.29 \times \sqrt{0.53 \times 10.34} \\ &= 21.68 \text{ kg}^2 \end{aligned}$$

Confidence intervals are constructed as

$$\hat{y} \pm t_{(\alpha/2)} \sqrt{\widehat{\text{var}}(\hat{y})}$$

For an approximate 95% confidence limit, we will use  $\pm 2(21.68 \text{ kg}^2)^{0.5}$  giving

$$[16] \quad 268.1 \pm 9.3 \text{ kg}$$

For a prediction interval we will need the variances computed from eq. 11. For  $\hat{y}_{\text{wood}(\text{new})}$  we obtain from Table 2 the values  $\hat{\sigma}^2 = 2.3446 \times 10^{-6} \text{ kg}^2$  and  $\psi(\hat{\theta}) = (D^2 H)^{2.114} = (20^2 \times 17)^{2.114} = 126 \, 451 \, 458$ . From the variance for  $\hat{y}_{\text{wood}}$  we know that  $\mathbf{z}(\mathbf{b})' \mathbf{s}^2(\mathbf{b}) \mathbf{z}(\mathbf{b}) = 8.46 \text{ kg}^2$ . Therefore, from eq. 11 we obtain

$$\widehat{\text{var}}(\hat{y}_{\text{wood}(\text{new})}) = 2.3446 \times 10^{-6} \text{ kg}^2 \times 126 \, 451 \, 458 \times 8.46 \text{ kg}^2 = 304.94 \text{ kg}^2$$

For the other components we obtain:  $\widehat{\text{var}}(\hat{y}_{\text{bark}(\text{new})}) = 18.66 \text{ kg}^2$  and  $\widehat{\text{var}}(\hat{y}_{\text{crown}(\text{new})}) = 23.192 \text{ kg}^2$ . From eq. 13 (and the previously computed correlations) we have

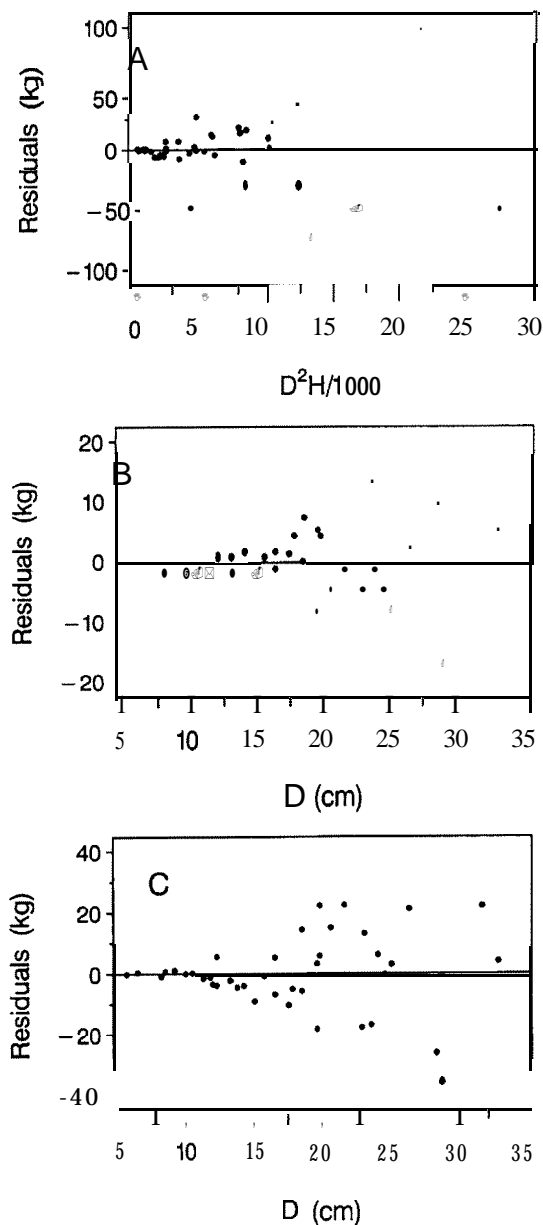
$$\begin{aligned} \widehat{\text{var}}(\hat{y}_{\text{total}(\text{new})}) &= 304.94 + 18.66 + 23.192 + 2 \times 0.19 \times \sqrt{304.94 \times 18.66} + 2 \times 0.01 \times \sqrt{304.94 \times 23.192} \\ &\quad + 2 \times 0.29 \times \sqrt{18.66 \times 23.192} = 627.66 \text{ kg}^2 \end{aligned}$$

Hence, an approximate 95% prediction interval on  $\hat{y}_{\text{total}(\text{new})}$  is (using  $t = 2$ ):

$$[17] \quad 268.1 \pm 50.1 \text{ kg}$$

Some readers may be surprised about how much larger the  $\widehat{\text{var}}(\hat{y}_{\text{total}(\text{new})})$  is over  $\widehat{\text{var}}(\hat{y}_{\text{total}})$  in this example. While the  $\widehat{\text{var}}(\hat{y}_{\text{total}(\text{new})})$  is expected to be larger because of the second variance component, heteroscedasticity reflected in the size of  $\psi(\hat{\theta})$  has a big influence on the overall variance estimate. Without correcting for heteroscedasticity, variance on the larger trees would be underestimated and variance on the smaller trees would be overestimated.

Fig. 1. Scatterplots of residuals from (A) wood regression, (H) bark regression, and (C) crown regression showing significant heteroscedasticity.  $D$ , diameter breast height;  $H$ , tree height.



## Examples using procedure 2: NSUR estimation

### Initial remarks

In this procedure we consider a set of nonlinear models whereby we allow statistical dependence among components and the total tree biomass. A contemporaneously correlated set of nonlinear biomass models whose parameters are estimated by NSUR with parameter restrictions will result in efficient estimates and additive predictions. The model for total biomass must be a combination of the component biomass models to be additive. Using the same equations listed in eq. 15 the system of equations becomes

$$\begin{aligned}
 \hat{y}_{\text{wood}} &= b_{11}(D^2H)^{b_{12}} \\
 \hat{y}_{\text{bark}} &= b_{21}D^{b_{22}} \\
 \hat{y}_{\text{crown}} &= b_{31}D^{b_{32}}H^{b_{33}} \\
 \hat{y}_{\text{total}} &= b_{11}(D^2H)^{b_{12}} + b_{21}D^{b_{22}} + b_{31}D^{b_{32}}H^{b_{33}}
 \end{aligned}
 \quad [18]$$

where  $\hat{y}_{\text{total}}$  is restricted to have the same independent variables and coefficients as the component equations. The same error or weight functions (Table 2) will be used for the 3 component models. The error function for  $\hat{y}_{\text{total}}$  was determined by modeling the residuals from  $\hat{y}_{\text{total}}$  using the (unweighted) coefficients from Table 2. From scatterplots and the AIC I chose the error model:  $\ln e_i^2 = \ln \sigma^2 + \theta \ln(D_i \times H_i) + v_i$ . Upon applying OLS, the following weight function resulted:

$$[19] \quad \psi(\hat{\theta}) = (D \times H)^{3.450}$$

A brief explanation of fitting a system of equations such as in eq. 18 by NSUR follows.

### NSUR estimation

The general case of  $M$  contemporaneously correlated nonlinear models can be written as

$$\begin{aligned}
 y_1 &= f_1(\mathbf{X}, \boldsymbol{\beta}) + \epsilon_1 \\
 [20] \quad y_2 &= f_2(\mathbf{X}, \boldsymbol{\beta}) + \epsilon_2 \\
 &\vdots \\
 y_M &= f_M(\mathbf{X}, \boldsymbol{\beta}) + \epsilon_M
 \end{aligned}$$

or alternatively

$$\mathbf{y} = \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\epsilon}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_M \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix}$$

The same matrix  $\mathbf{X}$  and the same parameter vector  $\boldsymbol{\beta}$  appear in all equations to allow for the possibility that some explanatory variables, and some parameters, could be common to more than one equation. Each equation can, of course, be a different nonlinear function of  $\mathbf{X}$  and  $\boldsymbol{\beta}$ . If weights are to be used, the system matrix of weights is written in block-diagonal form as

$$\boldsymbol{\Psi}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Psi}_1(\boldsymbol{\theta}_1) & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Psi}_2(\boldsymbol{\theta}_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Psi}_M(\boldsymbol{\theta}_M) \end{bmatrix}$$

where  $\boldsymbol{\Psi}_i$  (short for  $\boldsymbol{\Psi}_i(\boldsymbol{\theta}_i)$ ) is a diagonal matrix of weights for the  $i$ th system equation. Let us define a new matrix  $\mathbf{A}$  as

$$\mathbf{A} = \sqrt{\boldsymbol{\Psi}^{-1}}$$

where  $\sqrt{\boldsymbol{\Psi}^{-1}}$  is an elementwise square root operation on the inverse of the weight matrix (see Searle 1982, pp. 201-202). The implicit assumptions for eq. 20 are (i)  $E[\Delta_i \epsilon_i] = 0$ , the mean of the weighted residuals is zero for each system equa-

**Table 2.** Coefficients ( $\pm$ SE), and 95% confidence intervals (in parentheses), weight functions, and fit statistics from nonlinear regressions on slash pine biomass data.

	$b_1$	$b_2$	$b_3$	Weight function	$R^{2*}$	$\hat{\sigma}^2$
<b>Unweighted</b>						
Wood	0.016 542 $\pm$ 0.004 29 (0.008, 0.025)	1.0571 $\pm$ 0.0265 (1.003, 1.111)		1	0.984	752.15
Bark	0.055 098 $\pm$ 0.015 4 (0.024, 0.086)	2.1519 $\pm$ 0.0862 (1.977, 2.326)		1	0.962	25.56
Crown	0.012317 $\pm$ 0.007 69 (-0.003, 0.028)	3.0380 $\pm$ 0.310 (2.409, 3.667)	-0.33691 $\pm$ 0.321 (-0.987, 0.313)	1	0.917	160.80
<b>Weighted</b>						
Wood	0.016 363 $\pm$ 0.001 82 (0.013, 0.020)	1.0585 $\pm$ 0.0131 (1.032, 1.085)		$(D^2H)^{2.114}$	0.989	$2.3446 \times 10^{-6}$
Bark	0.046 277 $\pm$ 0.007 08 (0.032, 0.061)	2.2093 $\pm$ 0.0522 (2.104, 2.315)		$D^{3.715}$	0.976	$2.6609 \times 10^{-4}$
Crown	0.027 378 $\pm$ 0.012 8 (0.001, 0.053)	3.6804 $\pm$ 0.257 (3.159, 4.202)	-1.262 4 $\pm$ 0.375 (-2.021, -0.503)	$D^{7.322} e^{-0.006185H^2}$	0.927	$3.9417 \times 10^{-7}$

\* $R^2 = 1 - \text{SSE}/\text{SST}$ ; for weighted regression  $\text{SSE} = \hat{\mathbf{e}}' \Psi(\hat{\boldsymbol{\theta}})^{-1} \hat{\mathbf{e}}$  and  $\text{SST} = \tilde{\mathbf{y}}' \Psi(\hat{\boldsymbol{\theta}})^{-1} \tilde{\mathbf{y}}$ , where  $\tilde{\mathbf{y}} = \mathbf{y} - \bar{y}$  and  $\bar{y} = \text{vecdiag}(\Psi(\hat{\boldsymbol{\theta}})^{-1})' \mathbf{y} / \text{sum}(\text{vecdiag}(\Psi(\hat{\boldsymbol{\theta}})^{-1}))$  (Steel et al. 1996, pp. 284–285).

tion; and (2)  $E[\Delta \mathbf{e} \mathbf{e}' \Delta'] = \Sigma \otimes \mathbf{I}_T$ , where  $\otimes$  is the Kronecker product,  $\Sigma$  is an  $(M \times M)$  weighted covariance matrix whose  $(i, j)$ th element is given by  $\sigma_{ij}$  (the covariance of the errors from weighted equations  $i$  and  $j$ ) and  $E[\Delta_i \mathbf{e}_i \mathbf{e}_j' \Delta_j'] = \sigma_{ij} \mathbf{I}$ . To form the NSUR estimator we need  $\Sigma$ . Generally the variances and covariances of eq. 20 are unknown and must be estimated. To estimate the  $\sigma_{ij}$ , we first fit each system equation by EGNLS and obtain the residuals  $\mathbf{e}_i = \mathbf{y}_i - \mathbf{f}_i(\mathbf{X}, \mathbf{b})$ . Consistent estimates of the variances and covariances are then calculated by

$$[21] \quad \hat{\sigma}_{ij} = \frac{1}{(T - K_i)^{0.5} (T - K_j)^{0.5}} \mathbf{e}_i' \hat{\Delta}_i' \hat{\Delta}_j \mathbf{e}_j$$

where the degrees-of-freedom corrections  $K_i$  and  $K_j$  are the number of coefficients per equation (Greene 1999, p. 617). Let us define  $\hat{\Sigma}$  as the matrix containing the estimates  $\hat{\sigma}_{ij}$  from eq. 21. To specify the NSUR estimator, let us extend the notation from the Estimated generalized nonlinear estimation section. To be clear, there are  $T$  observations per equation,  $M$  equations, and  $K$  parameters ( $\boldsymbol{\beta}$  has dimension  $(K \times 1)$ ). As before, we need the matrix of partial derivatives of the residual with respect to the parameters. For our NSUR system the partial derivatives matrix  $\mathbf{F}(\boldsymbol{\beta})'$  is a  $(K \times MT)$  matrix given by

$$[22] \quad \mathbf{F}(\boldsymbol{\beta})' = \frac{\partial \mathbf{e}'}{\partial \boldsymbol{\beta}} = \left[ \frac{\partial \mathbf{f}_1'}{\partial \boldsymbol{\beta}}, \frac{\partial \mathbf{f}_2'}{\partial \boldsymbol{\beta}}, \dots, \frac{\partial \mathbf{f}_M'}{\partial \boldsymbol{\beta}} \right]$$

When evaluated at a particular value for  $\boldsymbol{\beta}$ , say  $\boldsymbol{\beta}_1$ , it will be written as  $\mathbf{F}(\boldsymbol{\beta}_1)$ . If each equation contains a different set of parameters,  $\partial \mathbf{e}' / \partial \boldsymbol{\beta}$  is a block-diagonal matrix. The generalized NSUR estimate of the vector  $\boldsymbol{\beta}$  is that value of  $\boldsymbol{\beta}$  that minimizes the residual sum of squares

$$[23] \quad R(\boldsymbol{\beta}) = \mathbf{e}' \Delta' (\Sigma^{-1} \otimes \mathbf{I}) \Delta \mathbf{e} = [\mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\beta})]' \Delta' (\Sigma^{-1} \otimes \mathbf{I}) \Delta [\mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\beta})]$$

Under the Gauss-Newton gradient minimization method, using the estimated weights and covariances, the estimated generalized NSUR iteration function is

$$[24] \quad \boldsymbol{\beta}_{n+1} = \boldsymbol{\beta}_n + l_n [\mathbf{F}(\boldsymbol{\beta}_n)' \hat{\Delta}' (\hat{\Sigma}^{-1} \otimes \mathbf{I}) \hat{\Delta} \mathbf{F}(\boldsymbol{\beta}_n)]^{-1} \mathbf{F}(\boldsymbol{\beta}_n)' \hat{\Delta}' (\hat{\Sigma}^{-1} \otimes \mathbf{I}) \hat{\Delta} [\mathbf{y} - \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}_n)]$$

where  $l_n$ , as in eq. 8, is the step length. When the process has converged we obtain the estimated generalized NSUR estimate  $\mathbf{b}$ . As with EGNLS, obtaining convergence can be a “fun” challenge. The analyst often has to change the step length, try new initial values, etc. It should be noted that one can take the NSUR process a step further by re-estimating  $\Sigma$  from the NSUR residuals (via eq. 21) and feeding this back into eq. 24 to form a new estimate for  $\boldsymbol{\beta}$ , and so on, until convergence. When the random errors follow a multivariate normal distribution this estimator will be the maximum likelihood estimator. For a discussion of the pros and cons of generalized least squares and maximum likelihood see Carroll and Ruppert (1988, pp. 18–23).

The estimated covariance matrix of the parameter estimates is calculated as

$$[25] \quad \hat{\Sigma}_{\mathbf{b}} = [\mathbf{F}(\mathbf{b})' \hat{\Delta}' (\hat{\Sigma}^{-1} \otimes \mathbf{I}) \hat{\Delta} \mathbf{F}(\mathbf{b})]^{-1}$$

The NSUR system variance is based on eq. 23 and is obtained from

$$[26] \quad \hat{\sigma}_{\text{NSUR}}^2 = \frac{R(\mathbf{b})}{MT - K} - \frac{\mathbf{e}' \hat{\Delta}' (\hat{\Sigma}^{-1} \otimes \mathbf{I}) \hat{\Delta} \mathbf{e}}{MT - K}$$



The estimated variance from the  $i$ th system equation on the  $t$ th observation  $\hat{y}_{it}$  (where for simplicity I drop the  $t$  subscript) is given by

$$[27] \quad S_{\hat{y}}^2 = \mathbf{f}_i(\mathbf{b})' \hat{\Sigma}_b \mathbf{f}_i(\mathbf{b})$$

where  $\mathbf{f}_i(\mathbf{b})'$  is a row vector for the  $i$ th equation from the partial derivatives matrix  $\mathbf{F}(\mathbf{b})$ . One can construct the biomass tables and the associated  $(1 - \alpha)$  confidence intervals for a mean value ( $\hat{y}_i$ ) and prediction intervals for a predicted new outcome ( $\hat{y}_{i(\text{new})}$ ) by the formulas for the biomass estimate from the  $i$ th system equation:

$$[28a] \quad \hat{y}_i = f_i(\mathbf{x}, \mathbf{b})$$

mean confidence interval:

$$[28b] \quad \hat{y} \pm t_{(\alpha/2)} \sqrt{S_{\hat{y}_i}^2}$$

and prediction interval:

$$[28c] \quad \hat{y}_{i(\text{new})} \pm t_{(\alpha/2)} \sqrt{S_{\hat{y}_i}^2 + \hat{\sigma}_{\text{NSUR}}^2 \hat{\sigma}_{ii} \psi_i(\hat{\theta}_i)}$$

where  $\hat{\sigma}_{ii} \psi_i(\hat{\theta}_i)$  is the estimate of the conditional variance of the  $i$ th system equation ( $\hat{\sigma}_{ii}$  is the  $(i, i)$ th element of  $\hat{\Sigma}$ , re. eq. 21, and adding the  $t$  subscript,  $\psi_i(\hat{\theta}_i)$  is the estimated weight).

#### Example 1

For an additive biomass system  $M = c + 1$ , thus the system under consideration in eq. 18 has  $A4 = 3 + 1 = 4$  equations. The specific system of four equations in eq. 18 can be combined into one model written in matrix notation as

$$[29] \quad \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \mathbf{y}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1(\mathbf{X}_1, \mathbf{0}, \mathbf{0}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) \\ \mathbf{f}_2(\mathbf{0}, \mathbf{X}_2, \mathbf{0}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) \\ \mathbf{f}_3(\mathbf{0}, \mathbf{0}, \mathbf{X}_3, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) \\ \mathbf{f}_4(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \boldsymbol{\epsilon}_3 \\ \boldsymbol{\epsilon}_4 \end{bmatrix}$$

or alternatively  $\mathbf{y} = \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\epsilon}$ , where the subscript 1 refers to the model for wood biomass; subscript 2 refers to the model for bark biomass; subscript 3 refers to the model for crown biomass; subscript 4 refers to the model for total tree biomass; and the vectors  $\mathbf{y}$ ,  $\boldsymbol{\beta}$ , and  $\boldsymbol{\epsilon}$  are stacked column vectors (in particular  $\boldsymbol{\beta} = [\boldsymbol{\beta}_1' \boldsymbol{\beta}_2' \boldsymbol{\beta}_3'] = [\beta_{11} \beta_{12} \beta_{21} \beta_{22} \beta_{31} \beta_{32} \beta_{33}]'$ ). The parameters in eq. 29 were estimated from eq. 24 (using PROC MODEL (SAS Institute Inc. 1993), and I supplied  $\hat{\Sigma}$ , calculated from eq. 21, to the MODEL procedure). Table 3 lists the coefficients and their standard errors for the weighted system. Note that the weighted NSUR standard errors for all seven coefficients are smaller than the corresponding EGNLS standard errors listed in Table 2. This indicates that there were significant contemporaneous correlations and a gain in efficiency in parameter estimation. Regression results, in terms of coefficients of determination and root mean square errors, are given in Table 4.

Using the coefficients in Table 3, if  $\mathbf{D} = 20$  cm,  $\mathbf{H} = 17$  m, and  $i = 2$  (bark biomass), we have from eq. 28a:  $\mathbf{x}' = [0 \ 20 \ 0 \ 0]$  and  $\hat{y}_2 = 0.046 \ 040 \ (2\hat{\sigma})^{2112} = 34.7$  kg. The error for the mean prediction is computed using eq. 27. For  $\hat{y}_2$  the vector  $\mathbf{f}_2(\mathbf{b})'$  is specified by

$$[0 \ 0 \ D^{b_{22}} \ b_{11} D^{b_{22}} \ \ln(D) \ 0 \ 0 \ 0]$$

thus

$$\mathbf{f}_2(\mathbf{b})' = [0 \ 0 \ 753.140 \ 32 \ 103.876 \ 68 \ 0 \ 0 \ 0]$$

The covariance matrix of  $\mathbf{b}$  is calculated from eq. 25 and is

$$\begin{bmatrix} 2.94 \times 10^{-6} & -2.12 \times 10^{-5} & 1.95 \times 10^{-6} & -1.42 \times 10^{-5} & -1.54 \times 10^{-6} & 1.83 \times 10^{-5} & 1.98 \times 10^{-5} \\ -2.12 \times 10^{-5} & 1.55 \times 10^{-4} & -1.36 \times 10^{-5} & 1.01 \times 10^{-5} & 1.14 \times 10^{-5} & -1.04 \times 10^{-4} & -1.82 \times 10^{-4} \\ 1.95 \times 10^{-6} & -1.36 \times 10^{-5} & 2.80 \times 10^{-5} & -2.07 \times 10^{-4} & -3.65 \times 10^{-6} & 1.70 \times 10^{-5} & 7.30 \times 10^{-5} \\ -1.42 \times 10^{-5} & 1.01 \times 10^{-4} & -2.07 \times 10^{-4} & 1.57 \times 10^{-3} & 3.00 \times 10^{-5} & -4.80 \times 10^{-5} & 6.98 \times 10^{-4} \\ -1.54 \times 10^{-6} & 1.14 \times 10^{-5} & -3.65 \times 10^{-6} & 3.00 \times 10^{-5} & 1.90 \times 10^{-5} & 8.20 \times 10^{-5} & -5.31 \times 10^{-4} \\ 1.83 \times 10^{-5} & -1.04 \times 10^{-4} & 1.70 \times 10^{-5} & -4.80 \times 10^{-5} & 8.20 \times 10^{-5} & 2.17 \times 10^{-2} & -2.42 \times 10^{-2} \\ 1.98 \times 10^{-5} & -1.82 \times 10^{-4} & 7.30 \times 10^{-5} & -6.98 \times 10^{-4} & -5.31 \times 10^{-4} & -2.42 \times 10^{-2} & 3.74 \times 10^{-2} \end{bmatrix}$$

So from eq. 27 we have

$$S_{\hat{y}_2}^2 = \mathbf{f}_2(\mathbf{b})' \hat{\Sigma}_{\mathbf{b}} \mathbf{f}_2(\mathbf{b}) = 0.44 \text{ kg}^2$$

A 95% confidence interval for this point estimate can be calculated from eq. 28b and is ( $t = 2$ )  $34.7 \pm 1.3$  kg. For  $i = 4$  (total tree biomass) we have  $\mathbf{x}' = [6800 \ 20 \ 20 \ 17]$  from eq. 28a and

$$\hat{y}_4 = b_{11}(6800)^{b_{12}} + b_{21}(20)^{b_{22}} + b_{31}(20)^{b_{32}}(17)^{b_{33}} = 264.2 \text{ kg}$$

For  $\hat{y}_4$  the vector  $\mathbf{f}_4(\mathbf{b})$  is specified by

$$\begin{bmatrix} (D^2 H)^{b_{12}} \\ b_{11}(D^2 H)^{b_{12}} \ln(D^2 H) \\ D^{b_{22}} \\ b_{21} D^{b_{22}} \ln(D) \\ D^{b_{32}} H^{b_{33}} \\ b_{31} D^{b_{32}} H^{b_{33}} \ln(D) \\ b_{31} D^{b_{32}} H^{b_{33}} \ln(H) \end{bmatrix}$$

The error is computed using eq. 27 and results in

$$S_{\hat{y}_2}^2 = \mathbf{f}_4(\mathbf{b})' \hat{\Sigma}_{\mathbf{b}} \mathbf{f}_4(\mathbf{b}) = 0.44 \text{ kg}^2$$

From eq. 28b the 95% confidence interval (using  $t = 2$ ) is

$$[30] \quad 264.2 \pm 6.3 \text{ kg}$$

In comparing the NSUR confidence interval in eq. 30 with the procedure 1 confidence interval in eq. 16, we see that the NSUR based confidence interval is 32% narrower ( $\pm 6.3$  vs.  $\pm 9.3$  kg).

The prediction limits for our point estimate of total tree biomass are calculated using eq. 28c. From eq. 26 we calculate the system variance:  $\hat{\sigma}_{\text{NSUR}}^2 = 0.933$ . From eq. 21 we calculate the variance of the fourth system equation:  $\hat{\sigma}_{44} = 8.46 \times 10^{-7} \text{ kg}^2$ . The weight factor for  $\hat{y}_{\text{total}}$  is computed from eq. 19:  $\Psi = (20 \times 17)^{3.45} = 541\,504\,742$ . Inserting the pieces of information into eq. 28c, we obtain the following 95% prediction interval:

$$[31] \quad 264.2 \text{ kg} \pm 2 \sqrt{9.77 \text{ kg}^2 + 0.93 \times 8.46 \times 10^{-7} \text{ kg}^2 \times 541504\,742} = 264.2 \pm 41.8 \text{ kg}$$

As is immediately obvious from comparing the prediction limits in eq. 31 with those in eq. 17, the procedure 2 or NSUR prediction limits are smaller, in fact 16.6% narrower. The interval from eq. 31 fits entirely inside the interval from eq. 17: [222.4, 306.01] versus [218.0, 318.21].

## Example 2

In the previous example we assumed that the error terms in our system were additive (as in eq. 1). Looking at the individual equations in eq. 18, one may want to assume equation errors that are multiplicative to derive log-linear models. The logarithmic transformation tends to stabilize heteroscedastic variance (if  $\sigma_{\epsilon}$  is proportional to  $E[y]$ ; Neter et al. 1985, pp. 137-138) and is an alternative to deriving weights for each equation in the system. For example with  $y_{\text{wood}}$  a model with a multiplicative error term and the corresponding log-linear model are

$$y_{\text{wood}} = \beta_{11}(D^2 H)^{\beta_{12}} \epsilon; \ln y_{\text{wood}} = \ln \beta_{11} + \beta_{12} \ln(D^2 H) + \ln \epsilon$$

Similarly,  $y_{\text{bark}}$  and  $y_{\text{crown}}$  can be linearized. However, because of the additivity restriction, the inherent model for  $y_{\text{total}}$  cannot be linearized. Thus NSUR must be used as opposed to linear seemingly unrelated regressions for the log-transformed equations in eq. 18. The resulting system of equations is

$$\begin{aligned} \ln \hat{y}_{\text{wood}} &= \ln b_{11} + b_{12} \ln(D^2 H) \\ [32] \quad \ln \hat{y}_{\text{bark}} &= \ln b_{21} + b_{22} \ln D \\ \ln \hat{y}_{\text{crown}} &= \ln b_{31} + b_{32} \ln D + b_{33} \ln H \\ \ln \hat{y}_{\text{total}} &= \ln[b_{11}(D^2 H)^{b_{12}} + b_{21} D^{b_{22}} + b_{31} D^{b_{32}} H^{b_{33}}] \end{aligned}$$

This system was fitted using NSUR (with PROC MODEL in SAS software; SAS Institute Inc. 1993). The coefficients and their SEs are listed in Table 3. The log-transformed NSUR coefficient SEs compare very favorably against the EGNLS

Table 3. Coefficients and standard errors ( $\pm$ SE) from fitting the slash pine biomass data with nonlinear seemingly unrelated regressions.

	Weighted	Log transformed
$b_{11}$	$0.016245 \pm 0.00172$	$0.016530 \pm 0.00170$
$b_{12}$	$1.0590 \pm 0.0124$	$1.0568 \pm 0.0121$
$b_{21}$	$0.046040 \pm 0.00530$	$0.046206 \pm 0.00559$
$b_{22}$	$2.2112 \pm 0.0397$	$2.2011 \pm 0.0423$
$b_{11}$	$0.014015 \pm 0.00436$	$0.035861 \pm 0.0143$
$b_{32}$	$3.5098 \pm 0.147$	$3.8704 \pm 0.135$
$b_{33}$	$-0.87190 \pm 0.193$	$-1.5717 \pm 0.216$

coefficient SEs in Table 2, being smaller for six of the seven coefficients. This is not surprising, since we know from example 1 there are contemporaneous correlations between the equations. In comparing the weighted NSUR coefficients and SEs against the log-transformed NSUR coefficients and SEs in Table 3, the major differences occur with  $b_{31}$  and  $b_{32}$ . These two coefficients essentially doubled in size under the log transform of the data and their SEs increased, threefold for  $b_{31}$ . This is indicative that  $\sigma_{\epsilon}$  is not strictly proportional to  $E[y]$ , especially for  $y_{\text{crown}}$ . Two fit statistics, fit index (FI, analogous to  $R^2$ ) and root mean square error (RMSE), for the equations in eq. 32 are given in Table 4. The values of FI and RMSE for the wood and bark equations between the weighted NSUR and log-transformed NSUR are essentially identical. The values of the fit statistics for the crown and tree equations are better using weighted NSUR. For this data, modeling the individual equation error structures gives comparatively better results than using the logarithmic transformation. This issue of using weighted versus log-transformed data will be considered more fully in the Discussion section.

Using the coefficients in Table 3 for log-transformed data, as before, if  $D = 20$  cm and  $H = 17$  m, we have

$$\ln \hat{y}_{\text{total}} = 5.5782;$$

$$\text{thus } \hat{y}_{\text{total}} = \exp(5.5782) = 264.6 \text{ kg}$$

To place prediction bounds on  $\ln \hat{y}_{\text{total}}$  use eq. 28c. The partial derivatives vector,  $f_4(\mathbf{b})$ , needed to compute  $S_{\hat{y}_i}^2$ , is specified as follows. Let  $u = b_{11}(D^2H)^{b_{12}} + b_{21}D^{b_{22}} + b_{31}D^{b_{32}}H^{b_{33}}$ , then we have

$$f_4(\mathbf{b}) = \begin{bmatrix} (D^2H)^{b_{12}}/u \\ b_{11}(D^2H)^{b_{12}} \ln(D^2H)/u \\ D^{b_{22}}/u \\ b_{21}D^{b_{22}} \ln(D)/u \\ D^{b_{32}}H^{b_{33}}/u \\ b_{31}D^{b_{32}}H^{b_{33}} \ln(D)/u \\ b_{31}D^{b_{32}}H^{b_{33}} \ln(H)/u \end{bmatrix}$$

For this example we have

$$S_{\hat{y}_i}^2 = 2.076 \times 10^{-4}, \\ \hat{\sigma}_{\text{NSUR}}^2 = 0.9884, \text{ and } \hat{\sigma}_{44} = 6.954 \times 10^{-3}$$

Table 4. Nonlinear seemingly unrelated regression results for the slash pine biomass system.

Model	Weighted		Log transformed	
	FI <sup>a</sup>	RMSE <sup>b</sup>	FI <sup>a</sup>	RMSE <sup>b</sup>
Wood	0.984	27.1	0.984	27.1
Bark	0.961	5.0	0.961	5.1
Crown	0.909	13.0	0.879	15.0
Tree	0.988	31.2	0.986	33.9

<sup>a</sup>Fit index;  $FI = 1 - \Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2$ , where the dependent variable  $y$  is in original data scale.

<sup>b</sup>RMSE is root mean square error in original data scale

Since no weight factor is used (the log transformation does the weighting) we set  $\psi = 1$ . An approximate 95% prediction interval ( $t = 2$ ) is

$$[33] \quad 5.5782 \pm 0.1683 = [5.4099, 5.74651]; \\ \text{on the arithmetic scale } [223.6, 313.11]$$

Note that the interval  $[223.6, 313.11]$  is not symmetric about  $\hat{y}_{\text{total}}$ . This occurs because the log transform is a nonlinear transformation. That aside, in comparing the interval in eq. 33 against the interval in eq. 31, we see that the log-transformed based interval encompasses an additional 5.9 kg width, which translates to a 7.1% increase.

## Discussion

Additivity of component biomass regression equations has concerned forestry professionals, since Kozak (1970) first presented methods of ensuring additivity with linear equations. With the speed afforded by today's computers and the availability of sophisticated software, nonlinear model development plays a more crucial role than ever before. However, care must be exercised to ensure reasonable functions. A critical issue in data analysis is the selection of a good approximating model that represents the data well. Depending on the number of variables available, graphs in two, three, and four dimensions (PROC SPECTRAVIEW in SAS (SAS Institute Inc. 1994) provides four-dimensional views of data by using a color gradient to display values of a fourth variable in three-dimensional space) can help in the selection of a number of empirical models. Theory can aid in selecting the independent variables and in specifying functional forms of the regression relation, such as geometry of the entity, knowledge of interaction effects, restrictions in the response space (e.g.,  $y > 0$ ), hypothetical data dependencies (e.g., lag effects, autocorrelation), and so on. A commonly used basis for choosing among competing models is Akaike's information criterion. The AIC is highly touted by a number of authors (e.g., Cavanaugh 1997; Burnham and Anderson 1998), as is Amemiya's (1985) prediction criterion (APC)<sup>5</sup>. A recent development that may prove to provide better model discrimination is the information complexity (ICOMP) criterion of Bozdogan (2000). The ICOMP criterion combines a badness-of-fit term with a measure of model complexity by

<sup>5</sup>The APC is computed as  $APC_i = \mathbf{e}_i' \mathbf{e}_i / (T - K_i) (1 + K_i / T)$ , where  $\mathbf{e}_i$  is the residual vector of the  $i$ th alternative equation,  $K_i$  is the number of coefficients, and  $T$  is the number of observations.

taking into account the interdependencies of the parameter estimates as well as the dependencies of the model residuals.

Unlike linear model estimation, nonlinear estimation can pose many challenges. For one, initial parameter estimates must be specified. A variety of methods are available for obtaining starting values. Often, experience can be utilized to provide good starting values. When it is possible, an initial consistent estimator of  $\beta$  will be a good starting vector. For example, many equation forms (ignoring error) can be linearized by a transformation (such as by logarithms) and undergo OLS estimation to obtain starting values, even though the underlying model is intrinsically nonlinear because of additive errors. This is how I obtained initial values in the Example 1 section above. Neter et al. (1985, p. 479) suggested the following procedure. Select  $K$  representative observations and set the nonlinear regression function  $f(\mathbf{x}_i, \beta)$  equal to  $y_i$  for each of the  $K$  observations (thereby ignoring the random error). Solve the  $K$  equations for the  $K$  parameters and use the solutions as the starting values. Still another possibility is to do a grid search in the parameter space. Select in a grid fashion various trial choices of  $\beta_0$ , evaluate the residual sum of squares  $S(\beta_0) = \sum_{i=1}^n [y_i - f(\mathbf{x}_i, \beta_0)]^2$  for each of these

choices, and use as the starting values the  $\beta_0$  vector for which  $S(\beta_0)$  is smallest.

Once the starting values for the parameters have been obtained, a modeler is still faced with many potential choices. An iteration method, whether Gauss-Newton, Marquardt, or one of the others, needs to be selected. A solution may be to a local minima, so several starting vectors should be tried to ensure that the global minimum of the residual sum of squares is obtained. Problems with convergence can arise for many reasons. The matrix of partial derivatives, the direction matrix, may be singular, possibly indicating an over-parameterized model. It is possible that parameters will enter a space where arguments to such functions as logs and square roots become illegal, resulting in overflows in computations. Obviously a careful choice of the model and starting values can circumvent such problems. Still, even with an appropriate model, the iteration method may lead to steps that do not improve the estimates. The modeler may need to control the step lengths,  $l_n$ , try a different set of starting values, or switch the iteration method. The gradient could be nearly flat causing small changes in the residual sum of squares and (or) small changes in the parameter estimates with successive iterations but still be far from the solution. Fine tuning of the convergence criteria may be necessary in such a situation to prevent premature stopping of the iterations and a subsequent loss of accuracy in the coefficients. As should be obvious, the selection and estimation of nonlinear models requires much more sophistication on the part of analysts than does linear models. Ross (1990) provides insights into why some models are difficult to fit and discusses the use of "stable parameter systems."

Heteroscedasticity is almost a certainty in biomass data and must be dealt with to achieve minimum variance estimates and reliable prediction intervals. The generalized method of moments (GMM) is an estimation method that produces efficient parameter estimates under heteroscedastic conditions without any specification of the nature of the heteroscedasticity (Greene 1999). However, one cannot generate bounds on the predictions

without specifying or otherwise estimating the error structure. Prediction intervals are necessary for reliability analysis and economic planning; therefore, I generally do not recommend GMM, preferring generalized least squares or the principle of maximum likelihood. The problem of heteroscedasticity then becomes one of estimating a weight factor or performing a transformation to stabilize variance. A treatise on the subject of transformation and weighting in regression is given by Carroll and Ruppert (1988). Theoretical considerations may lead one to use a transformation such as logarithms (e.g., if one believes the equation errors are multiplicative), but in practice I have found modeling the error structure gives empirical results as good as or better than applying transformations such as square roots or logarithms. In looking at Tables 3 and 4, the weighted and log-transformed equations for wood and bark biomass have similar SEs on the corresponding coefficients and closely matched FI and RMSE values; however, for the crown and total biomass regressions the weighted fit is superior in terms of SEs on the coefficients, FI, and RMSE. The log transformation is quick and simple to apply compared with the effort needed to develop error functions. The slight 7.1% increase in the log-transform based prediction interval (eq. 31 vs. 33) for the example total tree biomass prediction may seem inconsequential, but when considering hundreds or thousands of trees the cumulative effect can be considerable. Transformations such as the log or square root lead to predictions with an inherent bias on the original arithmetic scale. There are theoretical corrections for the bias (Miller 1984), but data rarely are "ideal"; thus, some studies have shown these corrections tend to overestimate the true bias (e.g., Madgwick and Satoo 1975; Hepp and Brister 1982). For these and other reasons I prefer to model the error structure rather than use the logarithmic transformation to correct for heteroscedasticity in biomass data. See Parresol (1993, 1999), Williams and Gregoire (1993), and Schreuder and Williams (1998) for examples on modeling the variance structure.

Iterative reweighted (nonlinear) least squares could be used as a further refinement to remedy heteroscedasticity. That is, start with the unweighted least squares estimate of  $\beta$ , compute residuals and fit the weight function then solve for  $\beta$ , compute residuals and refit the weight function then solve for  $\beta$ , and so on for  $C$  cycles. There is no clear consensus in the statistical literature about the best choice of the number of cycles, C. Goldberger (1964) and Matloff et al. (1984) propose  $C = 1$ , the latter showing from their simulation study that "the first iteration is usually the best, with accuracy actually deteriorating as the number of cycles is increased." Carroll and Ruppert (1988), based on their own simulation work, recommend  $C = 2$ . In my work with tree biomass data I have found little difference between using one or two cycles.

The property of additivity, as mentioned in the Introduction, leads to consistency among equations. The examples presented show a total tree (above stump) decomposed into three components: bole wood, bole bark, and tree crown. If the crown component were separated into two sub-components, say foliage and branches, we would want the regressions on the sub-components to give predictions that sum to the prediction from the crown regression, while still maintaining the overall additivity to the total. To maintain the property of additivity under this scenario using the

NSUR procedure requires that the crown component model be a combination of the subcomponent models with parameter restrictions between the subcomponent models and the component model. To illustrate, let us expand upon the system in eq. 18 by adding equations for foliage and branches:

$$\begin{aligned}
 \hat{y}_{\text{wood}} &= b_{11}(D^2H)^{b_{12}} \\
 \hat{y}_{\text{bark}} &= b_{21}D^{b_{22}} \\
 [34] \quad \hat{y}_{\text{foliage}} &= b_{31}D^{b_{32}}H^{b_{33}} \\
 \hat{y}_{\text{branches}} &= b_{41}(D \times \text{LCL})^{b_{42}} \\
 \hat{y}_{\text{crown}} &= b_{31}D^{b_{32}}H^{b_{33}} + b_{41}(D \times \text{LCL})^{b_{42}} \\
 \hat{y}_{\text{total}} &= b_{11}(D^2H)^{b_{12}} + b_{21}D^{b_{22}} + b_{31}D^{b_{32}}H^{b_{33}} \\
 &\quad + b_{41}(D \times \text{LCL})^{b_{42}}
 \end{aligned}$$

The expressed restrictions in eq. 34 guarantee that the foliage and branch predictions will sum to exactly equal the crown prediction, plus the total prediction will still exactly equal the sum of the component predictions. With the right restrictions many layers can be added to a system in such a way as to maintain the property of additivity. This property can be imposed on any quantity that can be disaggregated into a system of logical components.

I have presented methodology on two alternative procedures that guarantee additivity in the nonlinear case. The methodology presented shows computation of standard errors on the parameter estimates and computation of standard errors for confidence and prediction intervals on the predicted values. This allows judgment of the significance of coefficients (see Table 2, coefficients  $b_1$  and  $b_3$  for the crown regression) and aids in the assessment of the reliability of the equation predictions. Comparing standard errors of the coefficients between the two procedures for the slash pine biomass system shows a gain in efficiency using procedure 2, the NSUR approach, over procedure 1, the simple combination approach. Since components are not independent there will be contemporaneous correlations. The strength of the correlations determines the gain in efficiency. The tighter confidence and prediction intervals on the total biomass yields in eqs. 30 and 31, as compared with eqs. 16 and 17, are a direct consequence of the more efficient (minimum variance) NSUR estimator. In general, the procedure 2 or NSUR procedure can be recommended over procedure 1. Users can modify the program to fit any desired system of biomass equations.

If one is faced with a small sample size and if contemporaneous correlations are small, then there is a loss in finite sample efficiency with estimated generalized NSUR because of the uncertain estimator  $\hat{\Sigma}$ . In this case, procedure 1 is best. In my biomass work, modeling many different species, I have never found contemporaneous correlations to be sufficiently small to cause concern. Therefore, I almost unilaterally recommend the joint-generalized least squares approach for both linear and nonlinear modeling of biomass equation systems.

## Acknowledgements

The author thanks João Paulo Carvalho of the Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal, for discussions on the topic of nonlinear biomass systems, which

led to the writing of this paper. Gratitude is extended to Henry McNab and Dr. Stan Zarnoch of the USDA Forest Service, Asheville, N.C., for their reviews of this work. The author also thanks the Associate Editor and the three anonymous reviewers who provided invaluable suggestions that improved the scope and clarity of this work.

## References

- Amemiya, T. 1985. Advanced econometrics. Harvard University Press, Cambridge, Mass.
- Baldwin, V.C., Jr. 1987. A summary of equations for predicting biomass of planted southern pines. *In* Estimating Tree Biomass Regressions and Their Error. Proceedings of the Workshop on Tree Biomass Regression Functions and their Contribution to the Error of Forest Inventory Estimates. *Compiled by* E.H. Wharton and T. Cunia. USDA For. Serv. Gen. Tech. Rep. NE-117. pp. 157-171.
- Bates, D.M., and Watts, D.G. 1988. Nonlinear regression analysis and its applications. John Wiley & Sons, New York.
- Borowiak, D.S. 1989. Model discrimination for nonlinear regression models. Marcel Dekker, New York.
- Bozdogan, H. 2000. Akaike's information criterion and recent developments in information complexity. *J. Math. Psychol.* **44**(1): 62-91.
- Briggs, R.D., Cunia, T., White, E.H., and Yawney, H.W. 1987. Estimating sample tree biomass by subsampling: some empirical results. *In* Estimating Tree Biomass Regressions and Their Error. Proceedings of the Workshop on Tree Biomass Regression Functions and their Contribution to the Error of Forest Inventory Estimates. *Compiled by* E.H. Wharton and T. Cunia. USDA For. Serv. Gen. Tech. Rep. NE-117. pp. 119-127.
- Burnham, K.P., and Anderson, D.A. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York.
- Carroll, R.J., and Ruppert, D. 1988. Transformation and weighting in regression. Chapman & Hall, New York.
- Cavanaugh, J.E. 1997. Unifying the derivations for the Akaike and corrected Akaike information criteria. *Stat. Prob. Lett.* **33**: 201-208.
- Chiyenda, S.S., and Kozak, A. 1984. Additivity of component biomass regression equations when the underlying model is linear. *Can. J. For. Res.* **14**: 441-446.
- Clark, A., III. 1987. Summary of biomass equations available for softwood and hardwood species in the southern United States. *In* Estimating Tree Biomass Regressions and Their Error. Proceedings of the Workshop on Tree Biomass Regression Functions and their Contribution to the Error of Forest Inventory Estimates. *Compiled by* E.H. Wharton and T. Cunia. USDA For. Serv. Gen. Tech. Rep. NE-117. pp. 173-188.
- Cunia, T., and Briggs, R.D. 1984. Forcing additivity of biomass tables: some empirical results. *Can. J. For. Res.* **14**: 376-384.
- Cunia, T., and Briggs, R.D. 1985. Forcing additivity of biomass tables: use of the generalized least squares method. *Can. J. For. Res.* **15**: 23-28.
- Draper, N.R., and Smith, H. 1998. Applied regression analysis. 3rd ed. John Wiley & Sons, New York.
- Gallant, A.R. 1987. Nonlinear statistical models. John Wiley & Sons, New York.
- Goldberger, A.S. 1964. Econometric theory. John Wiley & Sons, New York.
- Greene, W.H. 1999. Econometric analysis. 4th ed. Prentice Hall, Upper Saddle River, N.J.

- Gregoire, T.G., Valentine, H.T., and Fumival, G.M. 1995. Sampling methods to estimate foliage and other characteristics of individual trees. *Ecology*, 76: 1181-1194.
- Harvey, A.C. 1976. Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, 44: 461-465.
- Hepp, T.E., and Brister, G.H. 1982. Estimating crown biomass in loblolly pine plantations in the Carolina flatwoods. *For. Sci.* 28: 115-127.
- Husch, B., Miller, C.I., and Beers, T.W. 1982. *Forest mensuration*. 3rd ed. John Wiley & Sons, New York.
- Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H., and Lee, T. 1985. *The theory and practice of econometrics*. 2nd ed. John Wiley & Sons, New York.
- Kleinn, C., and Pelz, D.R. 1987. Subsampling trees for biomass. In *Estimating Tree Biomass Regressions and Their Error. Proceedings of the Workshop on Tree Biomass Regression Functions and their Contribution to the Error of Forest Inventory Estimates*. Compiled by E.H. Wharton and T. Cunia. USDA For. Serv. Gen. Tech. Rep. NE-117. pp. 225-227.
- Kozak, A. 1970. Methods of ensuring additivity of biomass components by regression analysis. *For. Chron.* 46: 402-404.
- Lohrey, R.E. 1984. Aboveground biomass of planted and direct-seeded slash pine in the West Gulf Coast. In *Proceedings of the 6th Annual Southern Forest Biomass Workshop*, S-7 June 1984, Athens, Ga. Edited by J.R. Saucier. USDA Forest Service, Southeastern Forest Experiment Station, Asheville, N.C. pp. 75-82.
- Madgwick, H.A.I., and Satoo, T. 1975. On estimating the above ground weights of tree stands. *Ecology*, 56: 1446-1450.
- Matloff, N., Rose, R., and Tai, R. 1984. A comparison of two methods for estimating optimal weights in regression analysis. *J. Stat. Comput. Simul.* 19: 265-274.
- Miller, D.M. 1984. Reducing transformation bias in curve fitting. *Am. Stat.* 38: 124-126.
- Neter, J., Wasserman, W., and Kutner, M.H. 1985. *Applied linear statistical models*. 2nd ed. Irwin, Homewood, Ill.
- Pardé, J. 1980. Forest biomass. *For. Abstr.* 41: 343-362.
- Parresol, B.R. 1993. Modeling multiplicative error variance: an example predicting tree diameter from stump dimensions in baldcypress. *For. Sci.* 39: 670-679.
- Parresol, B.R. 1999. Assessing tree and stand biomass: a review with examples and critical comparisons. *For. Sci.* 45(4): 573-593.
- Pelz, D.R. 1987. Biomass studies in Europe-an overview. In *Estimating Tree Biomass Regressions and Their Error. Proceedings of the Workshop on Tree Biomass Regression Functions and their Contribution to the Error of Forest Inventory Estimates*. Compiled by E.H. Wharton and T. Cunia. USDA For. Serv. Gen. Tech. Rep. NE-117. pp. 213-224.
- Ratkowsky, D.A. 1983. *Nonlinear regression modeling: a unified practical approach*. Marcel Dekker, New York.
- Ross, G.J.S. 1990. *Non-linear estimation*. Springer-Verlag, New York.
- SAS Institute Inc. 1989. *SAS/STAT® user's guide*, version 6. 4th ed. Vol. 2. SAS Institute Inc., Cary, N.C.
- SAS Institute Inc. 1990. *SAS® procedures guide*, version 6. 3rd ed. SAS Institute Inc., Cary, N.C.
- SAS Institute Inc. 1993. *SAS/ETS® user's guide*, version 6. 2nd ed. SAS Institute Inc., Cary, N.C.
- SAS Institute Inc. 1994. *Highlighting SAS/SPECTRAVIEW™ software*, version 6. 1st ed. SAS Institute Inc., Cary, N.C.
- Schreuder, H.T., and Williams, M.S. 1998. Weighted linear regression using  $D^2H$  and  $D^2$  as the independent variables. USDA For. Serv. Res. Pap. RMRS-RP-6.
- Searle, S.R. 1982. *Matrix algebra useful for statistics*. John Wiley & Sons, New York.
- Seber, G.A.F., and Wild, C.J. 1989. *Nonlinear regression*. John Wiley & Sons, New York.
- Srivastava, V.K., and Giles, D.E.A. 1987. *Seemingly unrelated regression equations models: estimation and inference*. Marcell Dekker, New York.
- Steel, R.G.D., Dickey, D.A., and Torrie, J.H. 1996. *Principles and procedures of statistics: a biometrical approach*. 3rd ed. McGraw-Hill, New York.
- Valentine, H.T., Tritton, L.M., and Fumival, G.M. 1984. Subsampling trees for biomass, volume, or mineral content. *For. Sci.* 30: 673-681.
- Waring, R.H., and Running, S.W. 1998. *Forest ecosystems: analysis at multiple scales*. 2nd ed. Academic Press, San Diego, Calif.
- Williams, M.S., and Gregoire, T.G. 1993. Estimating weights when fitting linear regression models for tree volume. *Can. J. For. Res.* 23: 1725-1731.

## Erratum

Parresol, B.R. 2001. Additivity of nonlinear biomass equations. Can. J. For. Res. 31: 865-878.

Errors appear in the published version of this paper; the author requests that readers take note of the following corrections:

**p. 870** In the computation for  $\text{var}(\hat{y}_{\text{wood}(\text{new})})$ , it should read “+ 8.46 kg<sup>2</sup>” and not “× 8.46 kg<sup>2</sup>”.

**p. 871** Column 2, below eq. 20, the column vector for **e** should be

$$\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_M \end{bmatrix}$$

**p. 872** Within eq. 22 there should not be a prime symbol on the beta vector under  $\mathbf{f}_1$ , it should read  $\frac{\partial \mathbf{f}_1}{\partial \mathbf{b}}$ .

**p. 873** Equation 28b should start with  $\hat{y}_i$ , the subscript  $i$  was left off.

**p. 874** After the sentence “The error is computed using eq. 27 and results in” the formula should read  $S_{\hat{y}_4}^2 = \mathbf{f}_4(\mathbf{b})' \hat{\mathbf{S}}_{\mathbf{b}} \mathbf{f}_4(\mathbf{b}) = 9.77 \text{ kg}^2$ .

**p. 875** Column 1, first paragraph

“...the major differences occur with  $b_{31}$  and  $b_{32}$ .” Should read “... the major differences occur with  $b_{31}$  and  $b_{33}$ .”